Title
Complete genome sequence of the aerobic, heterotroph Marinithermus hydrothermalis type strain (T1(T)) from a deep-sea hydrothermal vent chimney.

Journal
Standards in genomic sciences, 6(1)

ISSN
1944-3277

Authors
Copeland, Alex
Gu, Wei
Yasawong, Montri
et al.

Publication Date
2012-03-01

DOI
10.4056/sigs.2435521

Peer reviewed

Complete genome sequence of the aerobic, heterotroph *Marinithermus hydrothermalis* type strain (T1^T) from a deep-sea hydrothermal vent chimney

Alex Copeland[1], Wei Gu[2], Montri Yasawong[3], Alla Lapidus[1], Susan Lucas[1], Shweta Deshpande[1], Ioanna Pagani[1], Roxanne Tapia[1,2], Jan-Fang Cheng[1], Lynne A. Goodwin[1,2], Sam Pitluck[1], Konstantinos Liolios[1], Natalia Ivanova[1], Konstantinos Mavromatis[1], Natalia Mikhailova[1], Amrita Pati[1], Amy Chen[4], Krishna Palaniappan[4], Miriam Land[1,5], Chongle Pan[1,5], Evelyne-Marie Brambilla[6], Manfred Rohde[7], Brian J. Tindall[6], Johannes Sikorski[6], Markus Göker[6], John C. Detter[1,2], James Bristow[1], Jonathan A. Eisen[1,8], Victor Markowitz[4], Philip Hugenholtz[1,9], Nikos C. Kyrpides[1], Hans-Peter Klenk[6*], and Tanja Woyke[1]

[1] DOE Joint Genome Institute, Walnut Creek, California, USA
[2] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
[3] Department of Biochemistry, Srinakharinwirot University, Bangkok, Thailand
[4] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[5] Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[6] Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[7] HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
[8] University of California Davis Genome Center, Davis, California, USA
[9] Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

***Corresponding author**: Hans-Peter Klenk (hpk@dsmz.de)

Keywords: strictly aerobic, non-motile, thermophilic, neutrophilic heterotroph, Gram-negative, hydrothermal vent, *Thermaceae*, GEBA

Marinithermus hydrothermalis Sako *et al*. 2003 is the type species of the monotypic genus *Marinithermus*. *M. hydrothermalis* T1^T was the first isolate within the phylum "*Thermus-Deinococcus*" to exhibit optimal growth under a salinity equivalent to that of sea water and to have an absolute requirement for NaCl for growth. *M. hydrothermalis* T1^T is of interest because it may provide a new insight into the ecological significance of the aerobic, thermophilic decomposers in the circulation of organic compounds in deep-sea hydrothermal vent ecosystems. This is the first completed genome sequence of a member of the genus *Marinithermus* and the seventh sequence from the family *Thermaceae*. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 2,269,167 bp long genome with its 2,251 protein-coding and 59 RNA genes is a part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project.

Introduction

Strain T1^T (= DSM 14884 = JCM 11576) is the type strain of the species *M. hydrothermalis*, which is the type species of the monotypic genus *Marinithermus* [1,2]. The genus name is derived from the Latin word 'marinus' meaning 'of the sea' and the latinized Greek word 'thermos' meaning 'hot', yielding the Neo-Latin word 'Marinithermus' meaning 'an organism living in hot marine places' [1]. The species epithet is derived from the Neo-Latin word 'hydrothermalis' (pertaining to a hydrothermal vent) [1]. Strain T1^T was isolated in November 2000 from the surface zone of a deep-sea hydrothermal vent chimney at Suiyo Seamount in the Izu-Bonin Arc, Japan, at a depth of 1,385 m [1]. *M. hydrothermalis* was the first isolate within the phylum "*Thermus-Deinococcus*" that grew optimally under a salinity equivalent to that of sea water [1].

The absolute requirement of NaCl for growth distinguishes *M. hydrothermalis* from members of the genera *Thermus* and *Meiothermus* [1,3]. No further isolates have been reported for *M. hydrothermalis*. Here we present a summary classification and a set of features for *M. hydrothermalis* T1^T, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *M. hydrothermalis* T1^T was compared using NCBI BLAST [4,5] under default settings (e.g., considering only the high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [6] and the relative frequencies of taxa and keywords (reduced to their stem [7]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Thermus* (91.0%), *Oceanithermus* (4.9%), *Marinithermus* (3.3%) and *Thermothrix* (0.8%) (118 hits in total). Regarding the two hits to sequences from members of the species, the average identity within HSPs was 100.0%, whereas the average coverage by HSPs was 98.0%. Among all other species, the one yielding the highest score was *O. profundus* (NR_027212), which corresponded to an identity of 91.9% and HSP coverage of 93.3%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was EU555123 [8] ('Microbial Sulfide Hydrothermal Vent Field Juan de Fuca Ridge Dudley hydrothermal vent clone 4132B16'), which showed an identity of 91.6% and HSP coverage of 92.1%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'spring' (6.9%), 'hot' (5.3%), 'microbi' (3.7%), 'nation, park, yellowston' (3.2%) and 'skin' (3.0%) (132 hits in total). Environmental samples which yielded hits of a higher score than the highest scoring species were not found. These key words are in accordance with the biotope of the strain T1^T in the original description [1], although 'skin' indicates the possible presence of relatives in a moderate environment.

Figure 1 shows the phylogenetic neighborhood of *M. hydrothermalis* T1^T in a 16S rRNA based tree. The sequences of the three identical 16S rRNA gene copies in the genome differ by two nucleotides from the previously published 16S rRNA sequence (AB079382).

The cells of strain T1^T are Gram-negative, non-motile, straight rods measuring 7.5 - 9.4 µm by 0.9 - 1.0 µm during the exponential growth phase [1] (Figure 2 and Table 1). In the stationary growth phase the cells tend to form filaments [1]. Rotund bodies were not observed from the cells [1]. Cells of strain T1^T have an envelope which consists of a cytoplasmic membrane with a simple outline and a cell wall with an inner, electron-dense thin layer, which presumably represents the peptidoglycan [1]. Colonies are whitish and have 2.5 - 3.0 mm of diameter [1]. The organism is an obligate heterotroph and grows only under strictly aerobic culture conditions [1]. Growth was not observed in anaerobic or autotrophic culture conditions [1]. However, it should be noted that according to Mori and colleagues [32] this was tested only in the presence of sulfide. Steinsbu and colleagues [3] argue that it is therefore possible that *M. hydrothermalis* has the capability of anaerobic growth under unreduced conditions, as has been observed for *Rhabdothermus arcticus*, *Vulcanithermus mediatlanticus*, *O. profundus* and *O. desulfurans* [3,32-34]. Unlike members of the genus *Thermus*, reactions were negative for catalase- and cytochrome oxidase and hydrolysis of gelatin, starch or casein was negative [1]. Growth occurs over the temperature range of 50.0 - 72.5°C (optimum 67.5°C), pH range 6.25 - 7.75 (optimum pH 7.0), and at NaCl concentrations in the range 0.5 - 4.5% (optimum 3%) [1]. The generation time under the above listed optimal condition and in medium MJYPV is about 30 minutes [1]. *M. hydrothermalis* T1^T differs from the members of the genera *Oceanithermus* by having a higher optimal temperature for growth and a higher oxygen tolerance [3]. Strain T1^T is able to utilize complex organic substrates such as Casamino acids, tryptone and yeast extract as sole energy and carbon sources [1].

Strain T1^T shares with its closest related genome-sequenced neighbors, *O. profundus* [17], *Meiothermus silvanus* [18] and *Thermus thermophilus* [16] (Figure 1), the presence of two linked 5S-23S rRNA gene clusters, with two 16S rRNA genes located separately in the genomes, but has one surplus, third 16S rRNA gene copy.

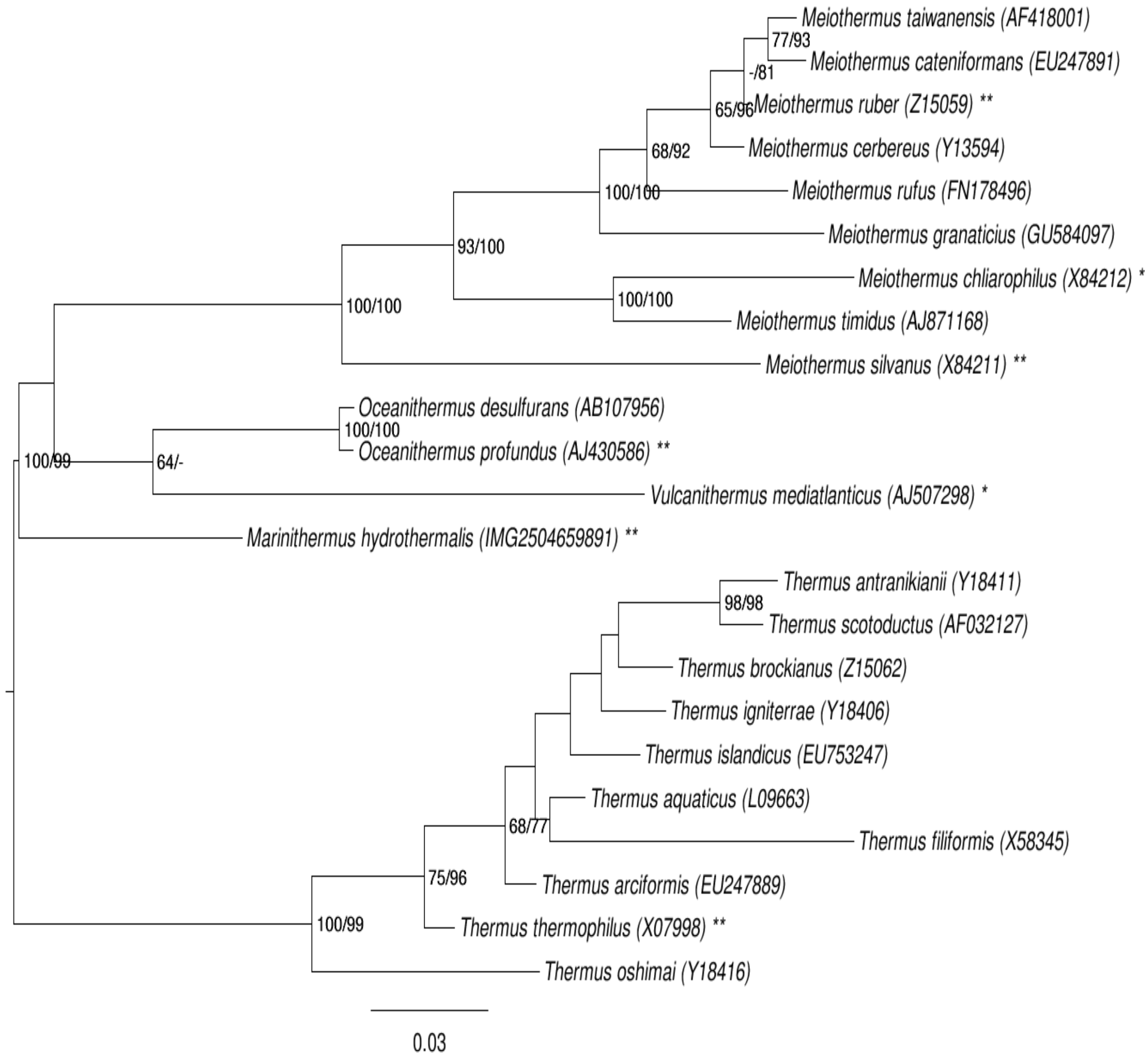

Figure 1. Phylogenetic tree highlighting the position of *M. hydrothermalis* relative to the type strains of the other species within the family *Thermaceae*. The tree was inferred from 1,426 aligned characters [9,10] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [11]. Rooting was done initially using the midpoint method [12] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 850 ML bootstrap replicates [13] (left) and from 1,000 maximum-parsimony bootstrap replicates [14] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [15] are labeled with an asterisk, those also listed as 'Complete and Published' with two asterisks [16-19].

Chemotaxonomy

The major cellular fatty acids of strain T1^T, when grown at 67.5°C, were *iso*-$C_{15:0}$ (40.4%), *iso*-$C_{17:0}$ (28.5%), $C_{16:0}$ (12.9%), *anteiso*-$C_{15:0}$ (6.0%), *anteiso*-$C_{17:0}$ (5.4%), *iso*-$C_{16:0}$ (2.8%) and *iso* 3-OH $C_{11:0}$ (1.0%). Menaquinone-8 was the major respiratory quinone. The fatty acid and respiratory quinone composition were similar to those of members of the genus *Thermus*, as described previously [35,36]. However, the presence of *iso* 3-OH $C_{11:0}$ in strain T1^T distinguishes it from *Thermus* species [1].

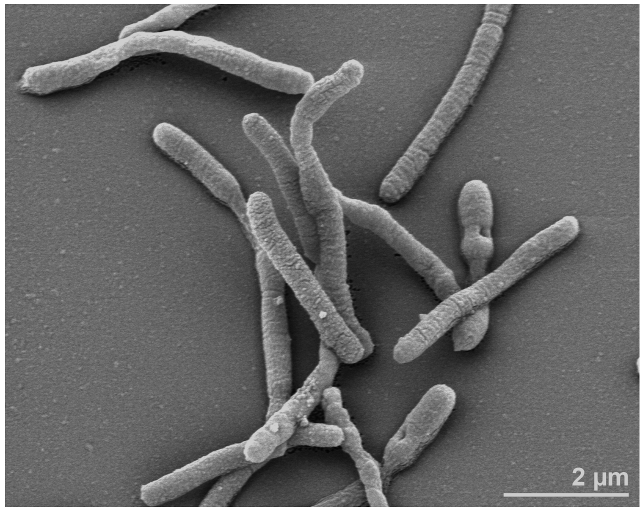

Figure 2. Scanning electron micrograph of *M. hydrothermalis* T1^{T}

Table 1. Classification and general features of *M. hydrothermalis* T1^{T} according to the MIGS recommendations [20] and the NamesforLife database [21].

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [22]
		Phylum *"Deinococcus-Thermus"*	TAS [23-25]
		Class *Deinococci*	TAS [26,27]
	Current classification	Order *Thermales*	TAS [26,28]
		Family *Thermaceae*	TAS [26,29]
		Genus *Marinithermus*	TAS [1]
		Species *Marinithermus hydrothermalis*	TAS [1]
		Type strain T1	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	straight rods	TAS [1]
	Motility	non-motile	TAS [1]
	Sporulation	none	NAS
	Temperature range	50.0°C-72.5°C	TAS [1]
	Optimum temperature	67.5°C	TAS [1]
	Salinity	0.5-4.5%, optimum 3% NaCl	TAS [1]
MIGS-22	Oxygen requirement	strictly aerobic	TAS [1]
	Carbon source	casamino acids, yeast extract, tryptone	TAS [1]
	Energy metabolism	neutrophilic heterotroph	TAS [1]
MIGS-6	Habitat	deep-sea, hydrothermal vent, marine	TAS [1]
MIGS-15	Biotic relationship	free-living	NAS
MIGS-14	Pathogenicity	not reported	
	Biosafety level	1	TAS [30]
	Isolation	deep-sea hydrothermal vent chimny	TAS [1]
MIGS-4	Geographic location	Suiyo Seamount, Izu-Bonin Arc, Japan	TAS [1]
MIGS-5	Sample collection time	November 2000	TAS [1]
MIGS-4.1	Latitude	28.65	TAS [1]
MIGS-4.2	Longitude	140.82	TAS [1]
MIGS-4.3	Depth	1,385 m	TAS [1]
MIGS-4.4	Altitude	- 1,385 m	TAS [1]

Evidence codes - NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [31].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [37], and is part of the ***G**enomic **E**ncyclopedia of **B**acteria and **A**rchaea* project [38]. The genome project is deposited in the Genomes OnLine Database [15] and the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

Growth conditions and DNA isolation

M. hydrothermalis T1^T, DSM 14884, was grown in DSMZ medium 973 (*Marinithermus hydrothermalis* medium) [39] at 70°C. DNA was isolated from 0.5-1 g of cell paste using MasterPure Gram-positive DNA purification kit (Epicentre MGP04100) following the standard protocol as recommended by the manufacturer, with modification st/DL for cell lysis as described in Wu *et al.* [38]. DNA is available through the DNA Bank Network [40].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [41]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 70 contigs in one scaffold was converted into a phrap [42] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (3,943.0 Mb) was assembled with Velvet [43] and the consensus sequences were shredded into 2.0 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 167.5 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [42] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [41], Dupfinisher [44], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 97 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [45]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 1,666.5 × coverage of the genome. The final assembly contained 458,684 pyrosequence and 48,027,166 Illumina reads.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Four genomic libraries: one 454 pyrosequence standard library, two 454 PE libraries (7.0 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	1,608.4 × Illumina; 58.1 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, Velvet version 0.7.63, phrap version SPS - 4.24
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002630
	Genbank Date of Release	April 15, 2011
	GOLD ID	Gc001721
	NCBI project ID	50827
	Database: IMG-GEBA	2504643006
MIGS-13	Source material identifier	DSM 14884
	Project relevance	Tree of Life, GEBA

Genome annotation

Genes were identified using Prodigal [46] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [47]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [48].

Genome properties

The genome consists of a 2,269,167 bp long chromosome with a 68.1% GC content (Figure 3 and Table 3). Of the 2,310 genes predicted, 2,251 were protein-coding genes, and 59 RNAs; 46 pseudogenes were also identified. The majority of the protein-coding genes (75.5%) were assigned with a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

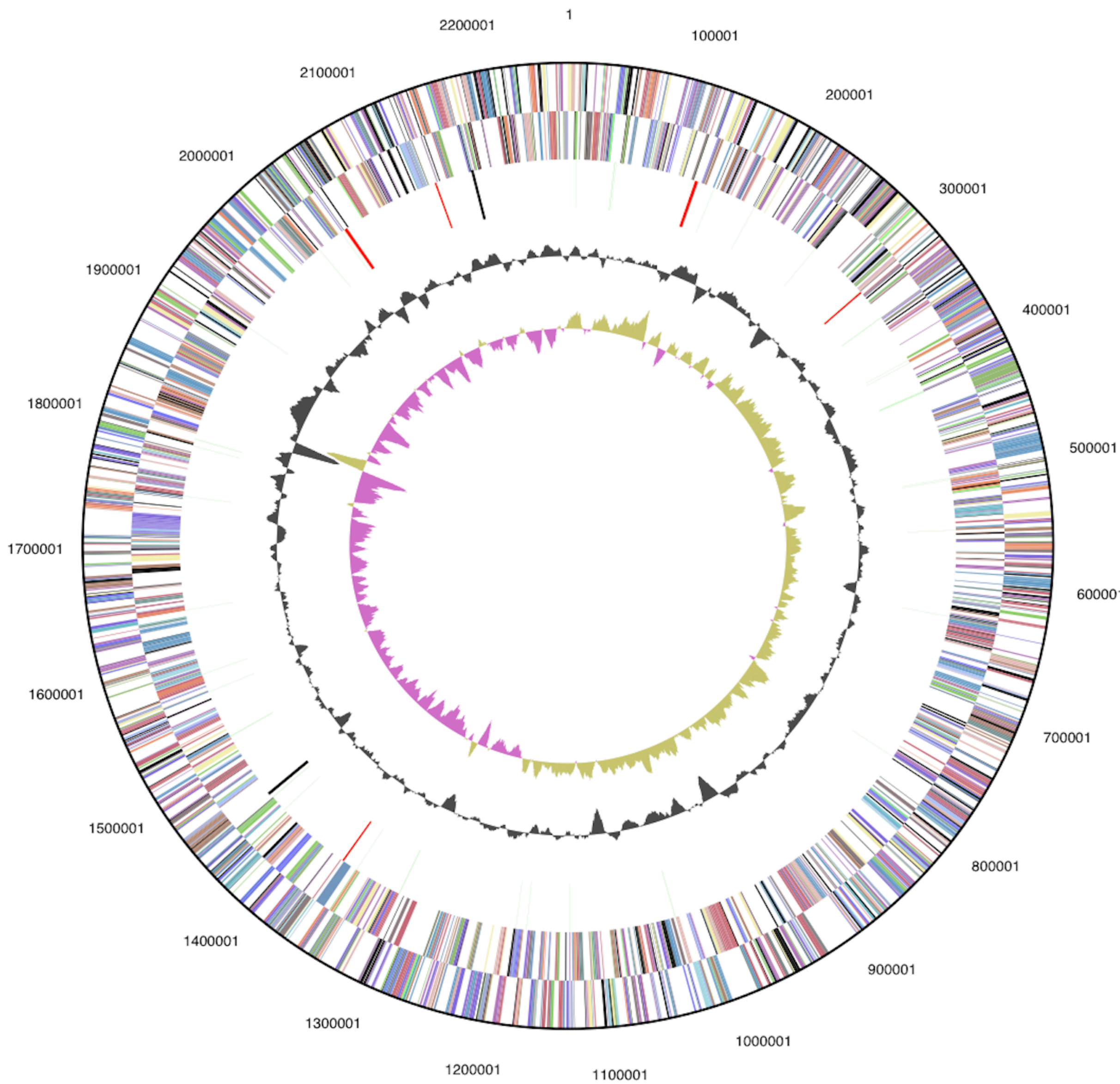

Figure 3. Graphical circular map of the chromosome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	2,269,167	100.00%
DNA coding region (bp)	2,092,686	92.22%
DNA G+C content (bp)	1,544,754	68.08%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	2,310	100.00%
RNA genes	59	2.55%
RNA operons	2*	
Protein-coding genes	2,251	97.45%
Pseudo genes	46	1.99%
Genes with function prediction	1,743	75.45%
Genes in paralog clusters	963	41.69%
Genes assigned to COGs	1,858	80.43%
Genes assigned Pfam domains	1,840	79.65%
Genes with signal peptides	479	20.74%
Genes with transmembrane helices	512	22.16%
CRISPR repeats	4	

* but three 16S rRNA genes

Table 4. Number of genes associated with the general COG functional categories

Code	Value	%age	Description
J	151	7.5	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	96	4.7	Transcription
L	99	4.9	Replication, recombination and repair
B	2	0.1	Chromatin structure and dynamics
D	27	1.3	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	30	1.5	Defense mechanisms
T	73	3.6	Signal transduction mechanisms
M	108	5.3	Cell wall/membrane/envelope biogenesis
N	21	1.0	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	49	2.4	Intracellular trafficking, secretion, and vesicular transport
O	86	4.2	Posttranslational modification, protein turnover, chaperones
C	149	7.4	Energy production and conversion
G	125	6.2	Carbohydrate transport and metabolism
E	215	10.6	Amino acid transport and metabolism
F	67	3.3	Nucleotide transport and metabolism
H	117	5.8	Coenzyme transport and metabolism
I	77	3.8	Lipid transport and metabolism
P	94	4.6	Inorganic ion transport and metabolism
Q	33	1.6	Secondary metabolites biosynthesis, transport and catabolism
R	255	12.6	General function prediction only
S	154	7.6	Function unknown
-	452	19.6	Not in COGs

Acknowledgements

We would like to gratefully acknowledge the help of Helga Pomrenke (DSMZ) for growing *M. hydrothermalis* cultures. This work was performed under the auspices of the US Department of Energy Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. Sako Y, Nakagawa S, Takai K, Horikoshi K. *Marinithermus hydrothermalis* gen. nov., sp. nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* 2003; **53**:59-65. PubMed doi:10.1099/ijs.0.02364-0
2. Euzeby JP. List of bacterial names with standing in nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**:590-592. PubMed doi:10.1099/00207713-47-2-590
3. Steinsbu BO, Tindall BJ, Torsvik VL, Thorseth IH, Daae FL, Pedersen RB. *Rhabdothermus arcticus* gen. nov., sp. nov., a novel member of the family *Thermaceae* isolated from a hydrothermal vent chimney from Soria Moria vent field at the Arctic Mid-Ocean Ridge. *Int J Syst Evol Microbiol* 2010. [Epub ahead of print, doi: 10.1099/ijs.0.027839-0].
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410. PubMed
5. Korf I, Yandell M, Bedell J. BLAST. Sebastopol, CA, O'Reilly, 2003.
6. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. PubMed doi:10.1128/AEM.03006-05
7. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.
8. Zhou H, Li J, Peng X, Meng J, Wang F, Ai Y. Microbial diversity of a sulfide black smoker in main endeavour hydrothermal vent field, Juan de Fuca Ridge. *J Microbiol* 2009; **47**:235-247. PubMed doi:10.1007/s12275-008-0311-z
9. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. PubMed
10. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. PubMed doi:10.1093/bioinformatics/18.3.452
11. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 2008; **57**:758-771. PubMed doi:10.1080/10635150802429642
12. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674. doi:10.1111/j.1095-8312.2007.00864.x
13. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? *Lect Notes Comput Sci* 2009; **5541**:184-200. doi:10.1007/978-3-642-02008-7_13
14. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). 4.0 b10. Sunderland, Sinauer Associates, 2002.
15. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. PubMed doi:10.1093/nar/gkp848
16. Henne A, Bruggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, Johann A, Lienard T, Gohl O, Martinez-Arias R, *et al*. The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat Biotechnol* 2004; **22**:547-553. PubMed doi:10.1038/nbt956
17. Pati A, Zhang X, Lapidus A, Nolan M, Lucas S, Del Rio TG, Tice H, Cheng JF, Tapia R, Han C, *et al*. Complete genome sequence of *Oceanithermus profundus* type strain (506^T). *Stand Genomic Sci* 2011; **4**:210-220. PubMed doi:10.4056/sigs.1734292
18. Sikorski J, Tindall BJ, Lowry S, Lucas S, Nolan M, Copeland A, Glavina Del Rio T, Tice H, Cheng JF, Han C, *et al*. Complete genome sequence of

Meiothermus silvanus type strain (VI-R2 ^{T}). *Stand Genomic Sci* 2010; **3**:37-46. PubMed doi:10.4056/sigs.1042812

19. Tindall BJ, Sikorski J, Lucas S, Goltsman E, Copeland A, Glavina Del Rio T, Nolan M, Tice H, Cheng JF, Han C, *et al*. Complete genome sequence of *Meiothermus ruber* type strain (21^{T}). *Stand Genomic Sci* 2010; **3**:26-36. PubMed doi:10.4056/sigs.1032748

20. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, *et al*. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed doi:10.1038/nbt1360

21. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.

22. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. PubMed doi:10.1073/pnas.87.12.4576

23. Garrity GM, Lilburn TG, Cole JR, Harrison SH, Euzéby J, Tindall BJ. Taxonomic outline of the Bacteria and Archaea, Release 7.7 March 6, 2007. Part 2 - The Bacteria: Phyla "*Aquificae*", "*Thermotogae*", "*Thermodesulfobacteria*", "*Deinococcus-Thermus*", "*Chrysiogenetes*", "*Chloroflexi*", "*Thermomicrobia*", "*Nitrospira*", "*Deferribacteres*", "*Cyanobacteria*", and "*Chlorobi*". http://www.taxonomicoutline.org/index.php/toba/article/view/187/211.

24. Weisburg WG, Giovannoni SJ, Woese CR. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst Appl Microbiol* 1989; **11**:128-134. PubMed

25. Weisburg WG, Giovannoni SJ, Woese CR. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst Appl Microbiol* 1989; **11**:128-134. PubMed

26. List Editor. Validation List no. 85. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol* 2002; **52**:685-690. PubMed doi:10.1099/ijs.0.02358-0

27. Garrity GM, Holt JG. Class I. *Deinococci* class. nov. *In*: Garrity GM, Boone DR, Castenholz RW, (*eds*), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 395.

28. Rainey FA, da Costa MS. Order II. *Thermales* ord. nov. *In*: Garrity GM, Boone DR, Castenholz RW, (*eds*), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 403.

29. da Costa MS, Rainey FA. Family I. *Thermaceae* fam. nov. *In*: Garrity GM, Boone DR, Castenholz RW, (*eds*), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 403-404.

30. BAuA. Classification of bacteria and archaea in risk groups. TRBA 466. Germany, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, 2010. p 125.

31. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. PubMed doi:10.1038/75556

32. Mori K, Kakegawa T, Higashi Y, Nakamura K, Maruyama A, Hanada S. *Oceanithermus desulfurans* sp. nov., a novel thermophilic, sulfur-reducing bacterium isolated from a sulfide chimney in Suiyo Seamount. *Int J Syst Evol Microbiol* 2004; **54**:1561-1566. PubMed doi:10.1099/ijs.0.02962-0

33. Miroshnichenko ML, L'Haridon S, Jeanthon C, Antipov AN, Kostrikina NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E, Bonch-Osmolovskaya EA. *Oceanithermus profundus* gen. nov., sp. nov., a thermophilic, microaerophilic, facultatively chemolithoheterotrophic bacterium from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 2003; **53**:747-752. PubMed doi:10.1099/ijs.0.02367-0

34. Miroshnichenko ML, L'Haridon S, Nercessian O, Antipov AN, Kostrikina NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E, Bonch-Osmolovskaya EA, *et al*. *Vulcanithermus mediatlanticus* gen. nov., sp. nov., a novel member of the family *Thermaceae* from a deep-sea hot vent. *Int J Syst Evol Microbiol* 2003; **53**:1143-1148. PubMed doi:10.1099/ijs.0.02579-0

35. Hensel R, Demharter W, Kandler O, Kroppenstedt RM, Stackebrandt E. Chemotaxonomic and molecular-genetic studies of the genus *Thermus*: evidence for a phylogenetic relationship of *Thermus aquaticus* and *Thermus ruber* to the

genus *Deinococcus*. *Int J Syst Bacteriol* 1986; **36**:444-453. doi:10.1099/00207713-36-3-444

36. Prado A, da Costa MS, Madeira VMC. Effect of growth temperature on the lipid composition of two strains of *Thermus* sp. *J Gen Microbiol* 1988; **134**:1653-1660.

37. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria? Syst Appl Microbiol* 2010; **33**:175-182. PubMed doi:10.1016/j.syapm.2010.03.003

38. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. *Nature* 2009; **462**:1056-1060. PubMed doi:10.1038/nature08656

39. List of growth media used at DSMZ: http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html.

40. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreservation and Biobanking* 2011; **9**:51-55. doi:10.1089/bio.2010.0029

41. DOE Joint Genome Institute. http://www.jgi.doe.gov.

42. Phrap and Phred for Windows. MacOS, Linux and Unix. http://www.phrap.com.

43. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. PubMed doi:10.1101/gr.074492.107

44. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. *In:* Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (*eds*), CSREA Press. June 26-29, 2006: 141-146.

45. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. 2008; Marco Island, FL.

46. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. PubMed doi:10.1186/1471-2105-11-119

47. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. PubMed doi:10.1038/nmeth.1457

48. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. PubMed doi:10.1093/bioinformatics/btp393